HXMT 高能望远镜毛刺的识别方法[1]

赵海升[1,2]，刘聪展[1,2]，李小波[1,2]，宋黎明[1,2]
1. 中国科学院粒子天体物理重点实验室，北京，100049；2. 中国科学院高能物理研究所，北京，100049

摘要：HXMT卫星高能望远镜数据存在非真实光子事例的聚集，突出表现在光变曲线上存在许多毛刺。这对光变和能谱的研究产生了一定的影响。本文分析了大量毛刺事例的性质，并基于这些性质通过事例时间间隔相邻法识别毛刺。我们发现这种方法可以在事例级别识别毛刺，这比利用一些特征函数，比如信噪比方法，有较大的优势。毛刺事例的识别对正常光子事例的影响非常小，而且毛刺事例在光变上不会产生周期和准周期信号。

关键词： 毛刺，时间间隔，信噪比，数据分析

中图分类号：p171.5

The spikes identification of of High Energy Telescope onboard Insight-HXMT

Zhao Haisheng, Liu Congzhan, Li Xiaobo, Song Liming

Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

Abstract: There are some spikes in the raw light curves of the High Energy Telescope (HE) onboard Hard X-ray Modulation Telescope (HXMT), and the appearance of these spikes has brought data analysis of HE with a lot of bad influence, especially for the study of light curves and energy spectra. A large amount of spikes are accumulated and these properties, such as time interval, duration time, channel distribution, are studied using statistical method. Base on these performances in light curves, two methods of spikes identification are established. In this work, we will introduce the properties and identification methods of spikes, and find the neighborhood method has more advantages than the other method.

Keywords: Spike, time interval, signal-to-noise, data analysis

0 引言

硬X射线调制望远镜（Hard X-ray Modulation Telescope，HXMT）卫星[1] 于2017年6月15日在酒泉发射升空，开启了中国在X射线波段对天体源的观测研究。它覆盖了1-250keV

1基金项目：国家自然科学基金（U1938102，11503029）.
作者：赵海升，男，博士. 研究方向：天文数据分析及算法. Email：zhaohs@ihep.ac.cn

能段，具有大天区和大有效面积巡天能力和高精度定点观测能力，同时还能够监测高能X射线和伽马射线暴发现象。HXMT搭载的高能望远镜（High Energy Telescope，HE）[2]是实现其科学探测能力的关键载荷。HE采用了条栅式准直器和NaI(TI)/CsI(Na)复合晶体探测器，由18个主探测器单元组成，NaI晶体可以对20~250keV的硬X射线实现高灵敏度观测，CsI晶体则可以对0.2-3MeV的暴发进行监测，同时两种晶体均具有能谱和时变观测能力。它的每个主探测单元由复合晶体探测器，光电倍增管（Photomultiplier Tube,PMT），前端电子学和后端电子学组成[3]。X射线光子入射到NaI晶体后将沉积全部能量，而CsI晶体做为主动式屏蔽探测器，对入射的粒子本底，伽马光子等进行监测，两种晶体产生的信号可通过脉冲形状甄别技术（脉冲宽度）加以区分。光电倍增管与晶体相连用于采集光信号，收集的光子信号经过前端电子学成形和放大，此时信号如果超过触发阈值，后端电子学将对信号进行采集，最终记录光子的能量，脉冲宽度和时间信息等。但是，如果信号超过上阈值（称之为大信号），后端电子学将不进行AD（Analog-to-Digital）转换而直接丢弃。HE的能量分辨率在60keV时接近19%，时间分辨率约8微秒，时间精度在2-4微秒，其快速的处理方式使得死时间仅占3~5%[2]。如此快速的数据处理也造成电子学产生一些非真实光子信号，并在两种晶体中同时存在，突出表现在光变曲线上出现一些尖峰，这些信号称为毛刺。

毛刺对光变和能谱的影响非常大。光变上，毛刺表现为一系列的尖峰，容易使人对光变的这种结构产生误解，比如可能认为这种现象是来自天体源的变化；在小天区扫描和伽马暴寻找方面，这些毛刺与源、伽马暴有一定程度的混淆，特别是对于一些能谱较软，流强较弱的短时伽马暴，比较难与毛刺区分；这些毛刺事例还会改变源光变的功率谱结构，在直接解调成像[4]中，它会引起本底（或者噪声）的涨落，对源寻找造成干扰。能谱上，它主要集中在低能段，对能谱拟合模型的选择和参数的约束带来一定的不确定性。当然，低能段事例可以被剔除，但是这样会降低源的显著性，毕竟低能段有效面积大，源光子数目多。这样剔除这些因为仪器原因引起的毛刺信号就变得非常重要。

目前，国内外对此现象的描述还比较少，本文将主要介绍毛刺的一些统计性质及识别方法，这对分析 HXMT 高能望远镜的数据有帮助，同时对未来的快速电子学处理系统的设计及后续的数据处理提供参考。另外，HE 探测器有 18 个单体，它们的电子学设计是一样的，各个单体产生的毛刺在分布和性质上也基本一致。本文主要以 2 号单体探测器为例进行描述。

1 毛刺的性质

毛刺的产生与晶体中大的能量沉积事例（大信号）有一定的关联。地面实验发现毛刺基本在大信号后产生，这些大信号可能引起 PMT 中空间电荷效应，进而引起大量的随机过阈事例。HE 在低增益观测时，其高压降低（使得同样能量的输出信号降低），毛刺将显著降低，这也说明毛刺与大信号的关系密切。

毛刺与 HE 电子学结构有比较大的关系，比如提高数据处理时间（增加死时间）或者降低高压，都可以显著的降低毛刺。然而毛刺的真实来源仍然不清楚，它的产生可能与晶体发光、光电子产生及前端电子学均有关系。目前地面实验不倾向于电子学噪声引起毛刺，而是倾向于大事例的输入引起，这样正常光子事例（20-250keV 光子）不认为能够产生毛刺。本文并不打算深究毛刺发生的根本原因，而是希望通过数据处理降低毛刺对正常事例的影响。本文对毛刺的识别也只能限于将显著的毛刺识别出来。

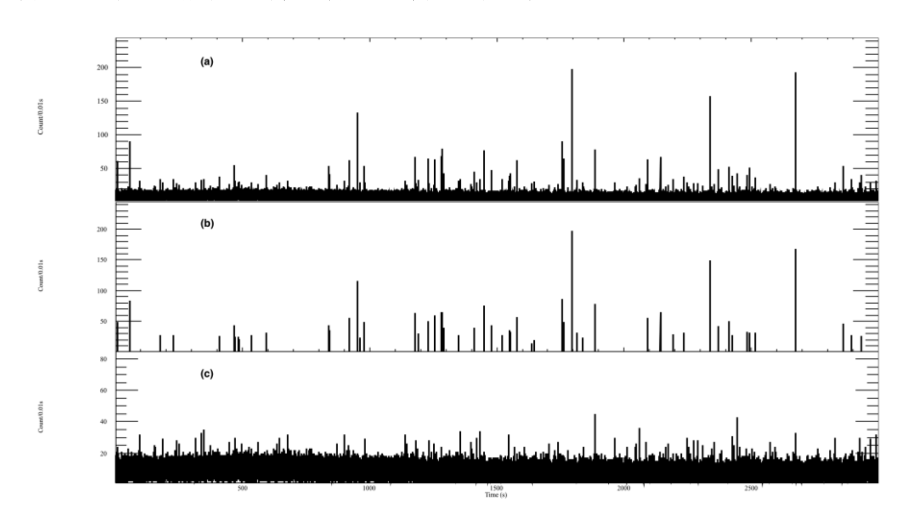


图 1 一段观测原始数据，挑选的毛刺和剔除毛刺后的光变，原始数据仅扣除了在轨放射源事例，单体为 2 号。(a)为原始的光变曲线，时间分辨为 0.01s；(b)为挑选的毛刺信号，时间分辨为 0.01s；(c)为剔除毛刺后的光变，时间分辨为 0.01s。

Fig. 1 a raw light curve of HE and only the calibration events are removed from this light curve: (a) the raw light curve with bin size 0.01 second; (b) the spikes with bin size 0.01 second; (c)the light curve removed spikes with bin size 0.01 second.

图 1(a)是一段 2 号单体探测器观测 Crab 源的原始数据的光变曲线，数据中仅扣除了在轨标定放射源事例，时间起点为 2018 年 11 月 20 日 13:40:26。源信号相对比较平稳，而本底主要是轨道粒子，在纬度比较高的区域计数会升高，而毛刺则在很短的时间内计数迅速升高并迅速下降。本文首先利用 HXMT 用户数据分析软件的方法（软件模块为 hepical）[5]，适当调整其默认参数得到图 1(b)的毛刺事例，我们可以清晰的看到该方法（将在下文介绍）将单个毛刺孤立的挑选出来；图 1(c)为剔除毛刺后的光变，其上面还有些疑似毛刺的事例。图 2 是利用 2 号单体探测器约 60ks 的 Crab 源观测数据统计毛刺的事例，大约得到 762 个孤立毛刺，事例数为 45807。(a)为这些事例在电子学的能道分布，HE 有自动增益控制，它可以根据在轨放射源峰位自动校正增益，所以采集这些光子的电子学增益是相同的，0-10 道为超上阈信号，这些事例主要由本底引起，毛刺事例则主要集中在 18-35 道，比较接近电子学阈值；(b)为脉冲宽度分布，该量与晶体发光时间有关，对于光子信号 80 以下主要是 NaI 事例，以上则为 CsI 事例，毛刺的脉冲宽度分布较广，几乎覆盖了 NaI 和 CsI 的典型宽

度；(c)为单个毛刺事例的时间间隔，多数毛刺事例的相邻时间小于 100 微秒，而且符合指数分布；(d)单个毛刺的持续时间分布，多数毛刺的持续时间在 1-2ms，但是该持续时间依赖于软件参数。

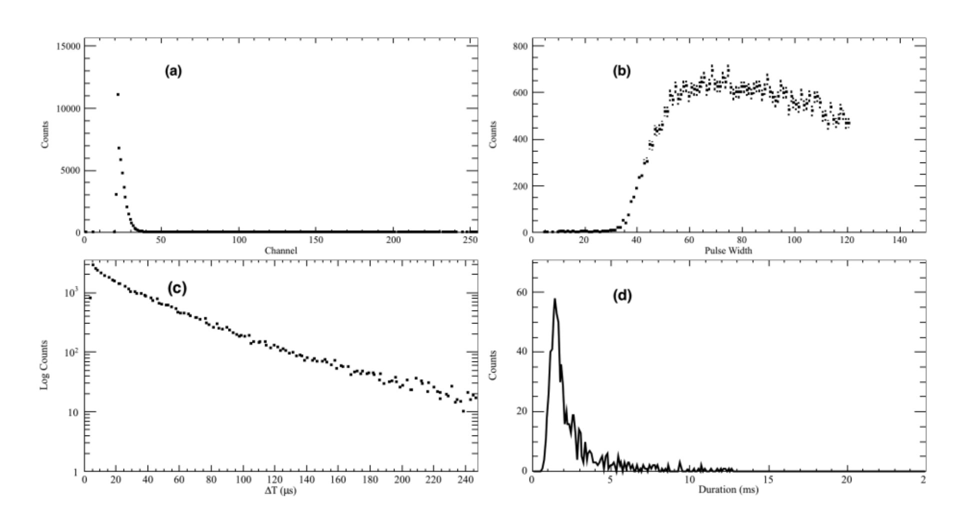


图 2 毛刺事例的分布：(a)为在电子学能道上的分布；(b)为脉冲宽度分布；(c)相邻事例的时间间隔；(d)毛刺持续时间分布。

Fig. 2 the properties of accumulated spikes: (a) channel distribution; (b) pulse width distribution; (c) time interval between continuous spike events; (d) duration time distribution.

2 毛刺的识别

毛刺的识别主要针对显著毛刺，即利用方法将短时间内的大量触发事例识别出来。如果将毛刺看为信号，而光子事例看为本底，则可以用信噪比判断毛刺（称为函数识别方法）。毛刺的持续时间多数在 5 毫秒内，故我们可以将光变的 bin 设置为 10 毫秒。本底的估计可以利用一段时间光变计数的平均或者线性拟合方法获得，可以适当将本底计数扩大（采用 1-3 倍标准偏差)以覆盖正常的光子计数区间，本底的计算步长可以选取 200 秒(或者更少)，这样本底计数可以看为基本不变。信噪比方法[6]采用 $SNR=S/(S+B)^{1/2}$, 图 3 为毛刺信噪比 SNR 大于 3 时挑选的毛刺，这个结果与图 1(b)是符合的。此方法存在两个参数，本底计数参数和 SNR 设置参数，本底计数参数可以根据实际测量给出，SNR 可以取 3 或者 5。SNR 公式也可以由其它方法取代，比如采用不完全伽马函数(IG(x,a))，它的形式如下：

$$IG(x,a) = \frac{1}{\Gamma(x)}\int_0^a t^{x-1}e^{-t}dt$$

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt\,, x > 0, a > 0.$$

a 代表本底，x 表示数据点，如果 IG(x,a)大于某一值(比如 0.999999)，则认为数据点为毛刺信号。

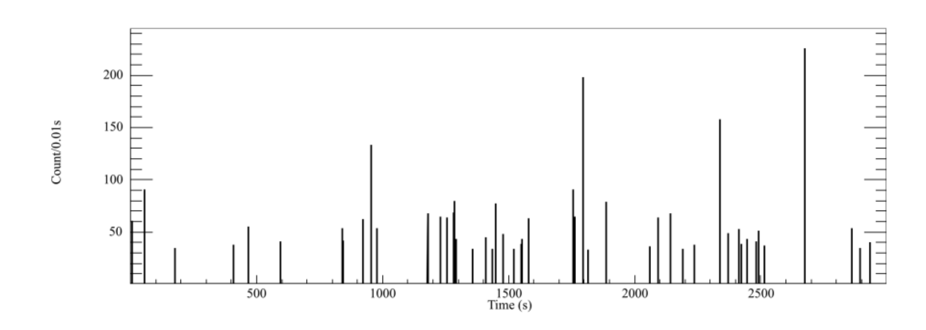


图 3 信噪比方法挑选的毛刺，数据中已经扣除在轨放射源事例，单体为 2 号，时间分辨为 0.01s。

Fig. 3 the light curve with bin size 0.01 second of spike events screened by the signal-to-noise method, and the data was acquired by detector of number 2 of HE.

目前 HXMT 的数据分析软件采用判断相邻事例时间间隔的方法识别毛刺（称为相邻法）：如果相邻事例的时间间隔小于 TIMEDEL，则使用计数器累计事例；如果持续有多个事例符合此条件，最终得到事例个数大于 EVTNUM，则认为这团事例为毛刺。图 1(b)是利用此方法得到的毛刺，TIMEDEL 和 EVTNUM 分别设置为 250 微秒和 25 个事例。这两个参数需要合理选择，一般来讲，TIMEDEL 越大，EVTNUM 也应该越大，这样做的好处是避免真正的光子事例被识别为毛刺事例；但是 EVTNUM 过大，则会造成一部分毛刺事例识别不出来。所以 EVTNUM 的选择有人为性。毛刺识别中没有对电子学能道和脉冲宽度进行限制，前者是防止在能谱上有截断，而后者是基于最大可能的利用毛刺计数高于信号（无论是 NaI 还是 CsI 信号，它们的计数都要小于毛刺计数）。

相邻法和函数识别法的区别在于前者是毛刺事例的识别，后者是毛刺时间 bin 的识别。两者需要设置参数的数目是一样的，均是 2 个参数，前者是事例时间间隔和事例总数，后者是本底数和判断函数的阈值。函数识别法，比如信噪比方法，参数的设置比较简单，而相邻法比较困难。如果毛刺来自电子学噪声随机过阈，这样正常光子事例也可以引起噪声变化，从而产生毛刺事例，一种解决办法是将相邻事例间隔比较小的事例扣除（可扣除能道小者，或者全部扣除），这样通过时间分 bin 是无法完成的，而在事例级别则可以（比如设置 EVTNUM 为 2，TIMEDEL 为 120 微秒）。另外，按 bin 识别毛刺，会造成毛刺事例分到相邻的 bin 上，容易漏掉一部分毛刺事例，而且正常事例也会被过滤掉，不过可以通过将 bin 值缩小而提高毛刺的识别，同时降低正常事例的扣除。总之，这两种方法在合适的参数下可以得到相似的结果，但是相邻法在毛刺识别中更有优势。

相邻法的参数为事例时间间隔（TIMEDEL）和事例总数（EVTNUM），本文在设置这两个参数时主要考虑到事例的时间间隔多在 250 微秒内，并且在毛刺的持续时间（毫秒内）中，可以认为本底和源的光子很少，如果毛刺显著的话，需要达到 25 个（$SNR=S/(S+B)^{1/2}=25/(25)^{1/2}=5$）。图 4（a）为将 TIMEDEL 和 EVTNUM 分别设置为 250 微秒和 5 个事例时，毛刺中光子事例数目的分布，随着毛刺事例数目的降低，毛刺个数也快速升高。

我们模拟了一段平均计数为 1000counts/s 的数据，利用相同的方法得到的毛刺事例数分布如图 5（a），识别出的毛刺的事例数截止到 10。这一结果依赖于平均计数的设置，而实际的 Crab 数据（源和本底）计数在 600-1400counts/s 之间，这样将 EVENUM 设置为 25，可以避免正常光子识别为毛刺。图 4（b）和图 5（b）为毛刺（或者识别出的毛刺）的持续时间分布，由毫秒（EVTNUM=25）缩减到几百微秒（EVTNUM=5）。

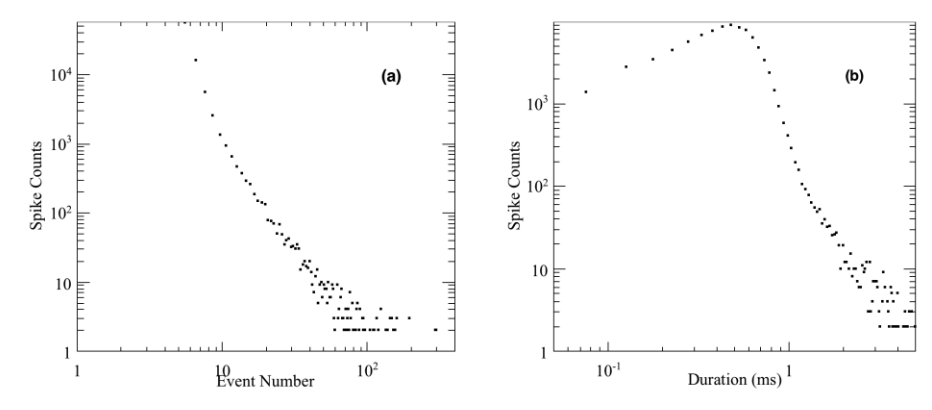


图 4 相邻法中将 TIMEDEL，EVTNUM 设置为 250 微秒和 5 个事例时，（a）毛刺中（假）光子事例数目的分布，其中横坐标为毛刺中光子事例的个数，纵坐标为毛刺计数；（b）毛刺持续时间分布，其中横坐标为毛刺的持续时间，纵坐标为毛刺计数。数据来自 Crab 数据，毛刺数目为 87104 个。

Fig. 4 the distributions of event number and duration time of spikes identified by the neighborhood method, when the parameters of TIMEDEL and EVTNUM are set to be 250 microsecond and 5 events, respectively. The data is originated from Crab observation and the total number of spikes is 87104.

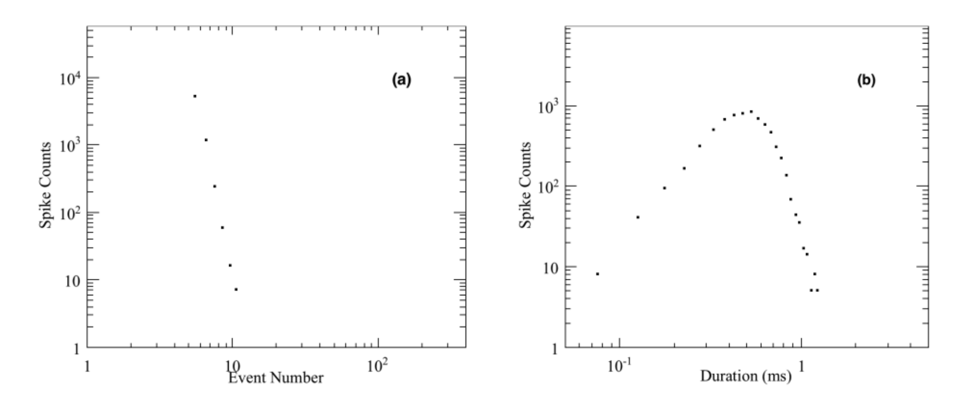


图 5 相邻法中将 TIMEDEL，EVTNUM 设置为 250 微秒和 5 个事例时，（a）识别出的毛刺事例数分布；（b）识别出的毛刺持续时间分布。数据来自产生 1000counts/s 的模拟数据，时间长度为 1 小时，识别出的毛刺数目为 6767 个。

Fig. 5 the distributions of event number and duration time of spikes identified by the neighborhood method, when the parameters of TIMEDEL and EVTNUM are set to be 250 microsecond and 5 events, respectively. The data is originated from Monte Carlo simulation and the counts rate is about 1000counts/s. The duration of data is 1 hour and the total number of identified spikes is 6767.

3 毛刺对数据分析的影响

HXMT 数据分析软件使用相邻方法，最大可能的识别毛刺事例，同时保留正常事例。但同时在事例筛选过后或多或少的对能谱和光变产生影响。

源能谱分析中，毛刺事例主要影响低能段能谱（电子学道数在 18-35），但是我们难以估计尚有多少毛刺事例未被识别，这是因为低能段的不确定因素比较多，比如探测器中晶体荧光产生效率的涨落和 PMT 光电子产生涨落使得光子探测效率难以有效确定，电子学噪声和粒子本底也使得本底难以有效估计，而且探测效率和本底的研究还有一定的耦合性。同时，各种方法不可避免的会将一部分源光子误识别为毛刺事例，这种误识别包括两种情况：(1)正常光子在毛刺发生时间内将会识别为毛刺事例；(2)正常光子事例与毛刺产生、结束时间接近，而识别为毛刺。相邻法产生的误识别主要是前者。我们选取一段 3000 秒时间的 Crab 数据，源和本底的光子计数率约为 800（counts/s），利用在轨标定放射源事例（计数率为 17counts/s，主要事例集中在电子学 50 道附近）估计误判率，在 52384 个放射源事例中，仅有 7 个识别为毛刺事例(约占毛刺事例的 0.2%)。相邻法对正常光子能谱也有一定的影响，这主要表现在曝光时间上，毛刺使得连续的观测数据出现毫秒和微秒的断裂，但是毛刺的持续时间多在 1-2 毫秒之间，对总曝光影响很少，比如在上述 3000 秒曝光时间中，毛刺总的持续时间仅占 0.11 秒。毛刺事例是否引起周期或者准周期信号，这是光变分析中需要解决的问题。图 6 是一段 Crab 观测数据中识别的毛刺事例的功率谱，从功率谱上没有看到明显的周期或者准周期信号。Crab 源存在大约 33 毫秒的周期信号，毛刺功率谱上没有看到此信号也说明正常事例识别为毛刺事例的比例非常少。

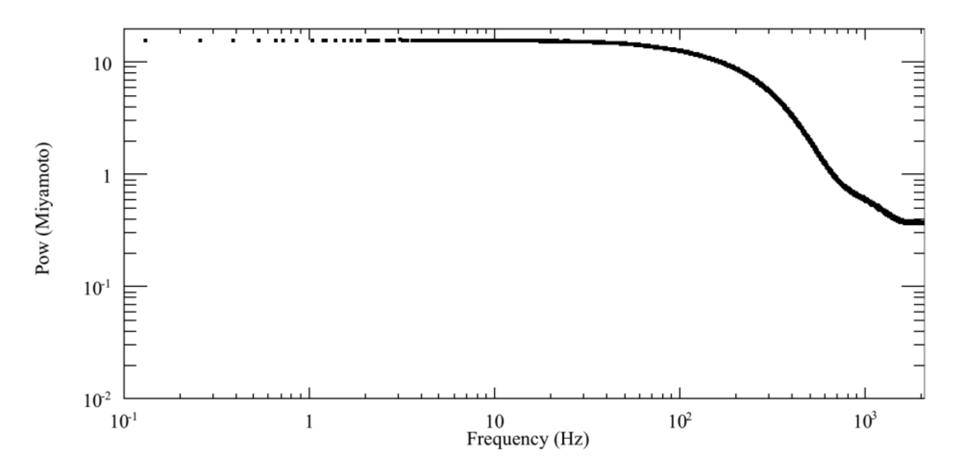


图 6 毛刺事例的功率谱，数据来自一段 Crab 数据中识别出的毛刺事例。毛刺事例非常不连续，对于不连续的光变 bin，填充了 0。

Fig. 6 power spectrum of spike events. The data is from Crab observation, and the bin without spike events will be filled 0.

4 总结及讨论

HXMT 卫星高能望远镜数据存在一些非真实光子事例的聚集，这在光变上表现为毛刺事例，它们的持续时间多在 2 毫秒内，事例时间间隔多在 100 微秒内，从能谱上看，毛刺事例主要分布在低能段（18-35 道），脉冲宽度分布较广，几乎覆盖了 NaI 和 CsI 的典型宽度。基于这些性质，我们采用判断相邻事例时间间隔的方法识别毛刺，这种方法可以在事例级别识别毛刺，这比利用一些特征函数，比如信噪比方法，有较大的优势。毛刺事例的识别对正常光子事例的影响非常小，而且毛刺事例在光变上不会产生周期和准周期信号。但是我们尚未真正理解毛刺产生的原因，只是发现与大信号有关系，而且我们也无法得知还有多少毛刺事例尚未识别出来。但是从光变上来讲，几乎所有显著的毛刺事例均识别出来。毛刺对低能段能谱的影响又与此能段标定，本底估计相互耦合。值得注意的是 HXMT 的本底估计依赖盲探测器的计数和能谱形状，而盲探测器与其它探测器在晶体结构和电子学上一致，未识别的毛刺可以通过对本底（盲探测器能谱）的扣除而压低。

本文主要在全能段扣除毛刺，实际上 HXMT 用户分析软件也提供了能段选择，可以在该能段中选择毛刺，而不影响其它能段。毛刺事例与真实光子产生的伽马暴等的最大区别是毛刺的持续时间非常短，而且不可能在多个单体上同时存在，我们可以利用这两个特点区分伽马暴和毛刺。

毛刺与大信号有关联，但是这些大信号的性质，比如与经纬度的关系，还有待分析。

参考文献

[1] Zhang, S.N., Li, T.P., Lu, F.J., Song, L.M., Xu, Y.P., Liu, C.Z., et al., 2020. Overview to the hard X-ray modulation telescope (Insight-HXMT) satellite. Sci. China, Phys. Mech. Astron. 63 (4), 249502. https://doi.org/10.1007/s11433-019-1432-6.

[2] Liu, C.Z., Zhang, Y.F., Li, X.F., Lu, X.F., Chang, Z., Zhang, A.M., et al., 2020. The high energy X-ray telescope (HE) onboard the Insight-HXMT astronomy satellite. Sci. China, Phys. Mech. Astron. 63 (4), 249503. arXiv:1910.04955.

[3] 李旭芳,刘聪展,张翼飞,李正伟,路雪峰,张爱梅,张钊,金永杰,硬X射线调制望远镜卫星高能望远镜设计与验证[J],航天器工程,2018年05期.

[4] Guan Ju; Lu Fangjun; Zhang Shu; et al. A modified direct demodulation method applied to Insight-HXMT Galactic plane scanning survey, Journal of High Energy Astrophysics, 2020, 26: 11-20.

[5] ZHAO Haisheng, LI Chengkui, LI Xiaobo, NIE Jianyin, GE Mingyu, PAN Yuanyue, SONG Liming. High Energy X-ray Telescope data analysis method[J]. Chin. J. Space Sci., 2016, 36(6): 938-944.

[6] ZHU Yongsheng. On Statistical Significance of Signal[J]. High Energy Physics and Nuclear Physics, 2006, 30(4).